

# Linkage between proton binding and folding in RNA: implications for RNA catalysis

P.C. Bevilacqua\*†<sup>1</sup>, T.S. Brown†, D. Chadalavada\*, J. Lecomte\*†, E. Moody\* and S.-i. Nakano‡

\*Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, U.S.A., †The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, U.S.A., and ‡Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 8-9-1 Okamoto, Higashinada-ku, Kobe 658-8501, Japan

## Abstract

Small ribozymes use their nucleobases to catalyse phosphodiester bond cleavage. The hepatitis delta virus ribozyme employs C75 as a general acid to protonate the 5'-bridging oxygen leaving group, and to accomplish this task efficiently, it shifts its  $pK_a$  towards neutrality. Simulations and thermodynamic experiments implicate linkage between folding and protonation in nucleobase  $pK_a$  shifting. Even small oligonucleotides are shown to fold in a highly co-operative manner, although they do so in a context-specific fashion. Linkage between protonation and co-operativity of folding may drive  $pK_a$  shifting and provide for enhanced function in RNA.

## Introduction

Catalytic RNAs, or ribozymes, were discovered almost 25 years ago [1,2]. Ribozymes can be loosely divided into large and small molecules. The first ribozymes to be discovered, group I introns and RNase P, are examples of large ribozymes (~300–400 nt), and this class cleaves phosphodiester bonds to afford products with 2',3'-hydroxy and 5'-phosphate termini. The smaller ribozymes (~40–80 nt), which include the hammerhead, hairpin, VS (Varkud satellite) and HDV (hepatitis delta virus) ribozymes, leave the opposite termini, a 2',3'-cyclic phosphate and a 5'-hydroxy group. Since their discovery, a large number of studies have been performed on ribozymes. Structures of many ribozymes have been solved by crystallographic and NMR studies, and the general nature of their reaction mechanisms has been elucidated (for reviews see [3–5], see also [6–9] and references therein). The largest ribozymes appear to act as metalloenzymes, using  $Mg^{2+}$  ions to stabilize unfavourable charge development in the transition state [6–9]. However, smaller ribozymes react at near-wild-type rates in the complete absence of divalent ions [10–13]. This phenomenon was first demonstrated for the hairpin, hammerhead and VS ribozymes by virtue of their reactivity in large quantities of EDTA at high pH [11]. Subsequently, this process was also shown for the HDV ribozyme, with low pH being required instead of high pH [12,13]. Apparently, the catalytic devices used by small ribozymes differ from those used by large ribozymes.

## The HDV ribozyme: an overview

Work in our laboratory has been engaged in investigating the catalytic mechanism of the HDV ribozyme. This ribozyme is approx. 85 nt in length and self-cleaves during rolling-

circle replication of the virus to afford linear monomers of the genome [14]. The secondary structures of the closely related genomic and antigenomic ribozymes were analysed in the 1990s, largely by mutagenesis and structure probing experiments in Been's laboratory [15], and revealed four pairings, numbered P1–P4 (Figure 1A). The crystal structure of the self-cleaved form of the ribozyme was later solved by Ferré-D'Amaré et al. [16] and revealed a compact fold, largely consistent with the secondary structure from Been and Wickham [15] (Figure 1B). The sole difference, a 2 base-pair pairing termed P1.1, was quickly shown to be relevant functionally in a collaboration between the two research groups [17]. Overall, the structure of the ribozyme is surprisingly complex, with two nested pseudo-knots and a buried active site. Of note is the close positioning of C75 and the 5'-hydroxy leaving group of G1 (Figure 1).

## The HDV ribozyme: mechanism

In the product structure, the N3 of C75 is only 2.7 Å ( $1 \text{ Å} = 10^{-10} \text{ m}$ ) from the O-5' of G1, suggesting a hydrogen bond (Figure 2A). Based on the principle of microscopic reversibility, this implicates N3 of C75 as the proton donor, or a general acid, in the cleavage reaction (Figure 2B). To serve as a general acid, the N3 must have a proton to donate. To have an appreciable population of protons to release at biological pH, the N3 must have a  $pK_a$  near the pH of the reaction, which is approx. 7.2 *in vivo*. The unshifted  $pK_a$  of C75 is approx. 4.2, requiring an upward shift of 3 units or approx. 4.2 kcal/mol ( $1 \text{ cal} = 4.184 \text{ J}$ ). Based on these considerations, we proposed that C75 might be a general acid in the reaction, offering a probable explanation for reactivity in the absence of divalent ions.

Experimental evidence for a  $pK_a$  of neutrality came from measuring the rate of self-cleavage as a function of pH. Experiments on the closely related antigenomic [18] and genomic [12] forms of the ribozyme showed  $pK_a$  values that varied

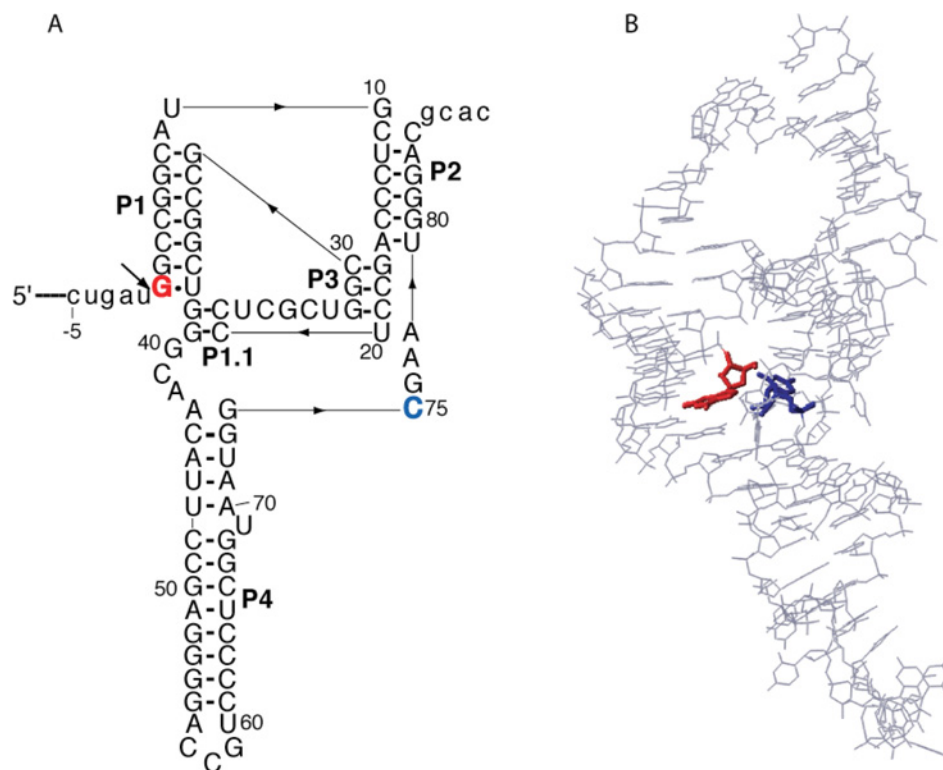
**Key words:** co-operativity, folding,  $pK_a$  shifting, proton binding, ribozyme, RNA catalysis.

**Abbreviations used:** HDV, hepatitis delta virus; VS, Varkud satellite.

<sup>1</sup>To whom correspondence should be addressed (email [pcb@chem.psu.edu](mailto:pcb@chem.psu.edu)).

**Figure 1 | Secondary and tertiary structures of the HDV ribozyme**

(A) Secondary structure of the HDV ribozyme. Five pairings (P1–P4 and P1.1) are denoted. The scissile phosphate is denoted with an arrow. G1 and C75 are shown in red and blue, respectively. (B) Tertiary structure of the cleaved form of the HDV ribozyme, solved in [16]. The Figure was made from pdb file 1dr3 using Web Lab Viewer Lite Software. Note the close positioning of G1 and C75.



from approx. 6 to >8, with higher values obtained at lower  $Mg^{2+}$  concentrations. Solvent isotope effects in the plateau region of the pH profile suggested that the rate-limiting step involved one or two proton transfers [12,19,20]. Site-directed mutants of the antigenomic and genomic ribozymes led to expected  $pK_a$  shifts, allowing the  $pK_a$  to be assigned to C75 [12,18]. However, kinetic ambiguity allowed the data to be interpreted in two different ways: C75 is the general acid or the general base in the cleavage reaction [21]. Scavenging of polyvalent ions by high concentrations of EDTA [performed in the presence of high concentrations of univalent ions (>100 mM) to promote tertiary structure] led to inversion of the pH profile of the reaction [12,13]. This observation along with positioning in the crystal structure helped to implicate C75 as the general acid in the cleavage reaction (Figure 2B).

It should be noted, however, that very recent crystal structures of the pre-cleaved form of the HDV ribozyme, inactivated by a C75U mutation, have led to a model supporting a general base role for C75 in cleavage [22]. Interpretation of these structures in terms of catalysis requires caution, however. Uracil lacks the exocyclic amine of cytosine, which engages in a hydrogen bond with the phosphate of C22 (Figure 2A), presumably important for positioning. In addition, when protonated at N3, uracil is neutral whereas cytosine is cationic. There is an extensive negative

potential near the active site in the product structure [13] and the reactive phosphate is negative, suggesting that a cationic moiety may be necessary to complete an electrostatic sandwich. Unexpectedly, a magnesium ion is observed near the leaving group oxygen in the C75U precursor structure, which led the authors to suggest that the hydrated  $Mg^{2+}$  serves as the general acid in the cleavage reaction [22]. Clearly, it will be important to solve a high-resolution structure of the pre-cleaved ribozyme with a cytosine at position 75.

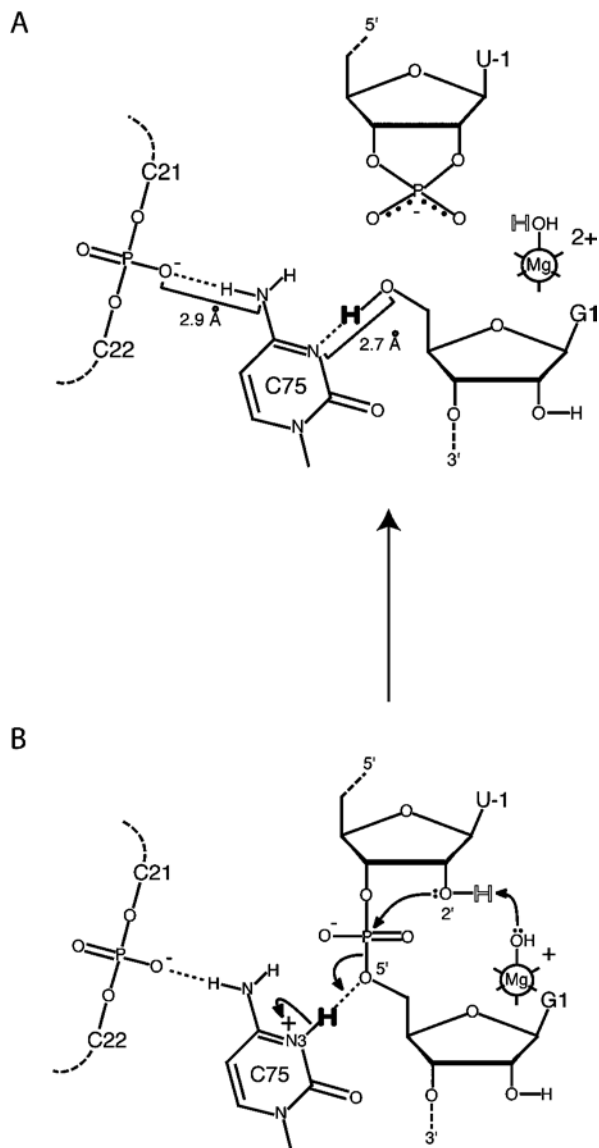
Rescue of inactive C75U antigenomic [18,20] and genomic [12] variants by exogenous nucleobase and imidazole derivatives further implicated C75 as a direct participant in the reaction. Subsequently, similar experiments have been performed on the hairpin ribozyme and support nucleobase involvement in the reaction [23], as do various crystal structures of this ribozyme [21,24,25]. It is becoming increasingly clear that the nucleobases themselves can participate in the making and breaking of covalent bonds, a role most closely analogous to histidines in the cleavage of RNA by the protein enzyme RNase A [26].

### The HDV ribozyme: a folding pathway

RNA molecules are prone to misfolding [27]. The four nucleobases can combine in alternative registers, employing

**Figure 2 | Proposed mechanism of the HDV ribozyme**

(A) Interactions noted in the crystal structure of the product form of the HDV ribozyme [16]. Note the close positioning of the N3 of C75 and the O5' of G1, as well as the N4 of C75 and the non-bridging phosphate oxygen of C22. (B) Pre-cleaved interactions inferred from the crystal structure and biochemical experiments. A hydrated magnesium hydroxide (not clearly observed in the product structure) serves as the general base, while protonated C75 serves as the general acid [12].



Watson–Crick and non-Watson–Crick base-pairs to form incorrect pairings. As the thermodynamic stability of RNA is high and since new helices cannot be made until alternative helices are either fully or partially broken, misfolding can lead to slow overall folding. Work in our laboratory has shown that the HDV ribozyme can adopt a number of misfolds involving both ribozyme–flanking and ribozyme–ribozyme pairings [28–31]. This work also led to the identification of upstream, P(-1), and downstream, P5, flanking-sequence–flanking-sequence pairings that aid folding by restricting

the availability of flanking sequences to interact with and misfold the ribozyme [28,29]. In addition, myriad non-native ribozyme–ribozyme pairings have been identified [30], including those that appear to facilitate ribozyme folding by preventing stronger alternative pairings from forming [31]. Impressively, many of these facilitating roles for alternative pairings were predicted from calculations [32].

One goal of these folding studies was the design of a sequence that maximizes the population of the native fold. Through these studies, we engineered a double mutant of the ribozyme that along with appropriate antisense oligonucleotides gives the fastest reacting HDV construct reported, cleaving in a largely monophasic fashion with an observed rate constant of 60 min<sup>-1</sup> [31]. Our studies and other recent studies on the VS [33] and hammerhead [34] ribozymes have demonstrated that small ribozymes have high intrinsic rates of chemistry (1–10 s<sup>-1</sup>) that are typically masked by high populations of non-native structures. Perhaps the conformational heterogeneity of RNA, rather than any intrinsic limitation to native state reactivity, hinders overall cleavage efficiency of ribozymes. It will be interesting to see the extent to which functionally relevant crystal structures of ribozymes resemble their protein counterparts.

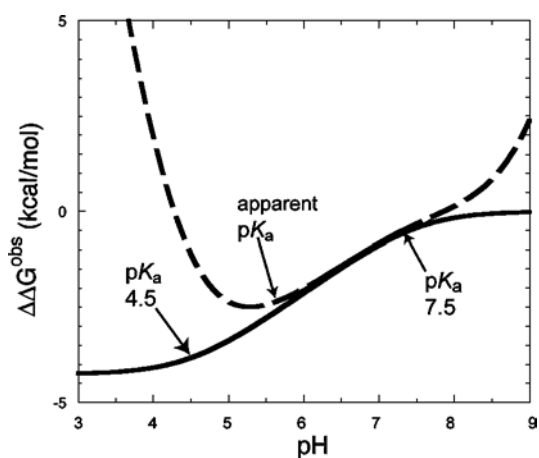
### Linkage between proton binding and folding

The observation that ribozymes can use their nucleobases in chemistry, raises the issue of what drives the p*K*<sub>a</sub> values of the bases towards neutrality? Folding and reactivity of the HDV ribozyme described in previous sections are intimately connected, and this linkage might be involved in p*K*<sub>a</sub> shifting. We have grouped p*K*<sub>a</sub> values into two classes based on whether the loaded proton is sequestered in hydrogen bonding (class I) or not (class II) [5]. Class II p*K*<sub>a</sub> values are obvious candidates for proton transfer, e.g. C75 in the HDV ribozyme, whereas class I p*K*<sub>a</sub> values may act as oxyanion holes as proposed for the ribosome [5]. Optimally, class II p*K*<sub>a</sub> values should be near 7 to strike a balance between being in the functional form and being a good proton donor/acceptor, while class I p*K*<sub>a</sub> values should be ≥8.5 to maximize the population of the cationic state. However, A and C residues in their unfolded state have p*K*<sub>a</sub> values of only approx. 3.7 and 4.4, respectively, for their imino nitrogens [35].

As shown by Misra and Draper [36], there is a linkage between Mg<sup>2+</sup> ion binding and RNA folding, wherein preferential binding of Mg<sup>2+</sup> ions to the folded state increases the stability of the folded structure. Recently, we performed a thermodynamic study of the linkage between proton binding and nucleic acid folding [35]. This was done in an effort to understand some of the driving forces for p*K*<sub>a</sub> shifting in RNA. It can be noted that these principles are more apt to apply to class I p*K*<sub>a</sub> values, whereas class II p*K*<sub>a</sub> values may be more influenced by electrostatics. Binding polynomials for the folded and unfolded states were enumerated and used to perform simulations of the dependence of free energy on pH (Figure 3). Among the features of these simulations,

### Figure 3 | Simulation of the dependence of net folding free energy on pH

The complex dependence of free energy on pH captures several features seen in experiments on model oligonucleotides, including acid and alkaline denaturation of the helix, a  $pK_a$  near neutrality, and an apparent unfolded state  $pK_a$  shifted towards neutral pH [35]. The broken line simulation is for a 24-mer oligonucleotide with a folded state  $pK_a$  of 7.5. The continuous line simulation assumes only one protonation in the unfolded state of  $pK_a$  4.5; this simulation serves to illustrate the importance of multiple protonations in the unfolded state.



there were steep increases in free energy at low and high pH. These increases arise because of the large number of proton-binding sites accessible in the unfolded state and are consistent with the well-known acid and alkaline denaturation of helices. Other features of the simulations were observation of a microscopic  $pK_a$  for the folded state, the feature of greatest interest to us, and an apparent  $pK_a$  for the unfolded state; this latter feature is statistical in nature and arises because of the many proton-binding sites available in the unfolded state. Melting experiments were performed on model oligonucleotides and these supported the thermodynamic formalism and provided experimental  $pK_a$  values in accordance with those determined independently.

One practical implication of these studies is that the extent of  $pK_a$  shifting should depend on the extent of folding interactions made possible by protonation. This was demonstrated directly by  $pK_a$  shifting of cationic AC wobbles to values of 7 in the presence of optimal nearest-neighbours [37]. Another descriptor of the extent of folding upon protonation is co-operativity. If folding is co-operative upon protonation, greater  $pK_a$  shifting should occur. Recently, we have made advances in understanding the co-operativity in RNA and DNA folding [38–40]. Folding of a DNA triloop and a related tetraloop with a minimal complement of interactions was shown to obey indirect coupling and be highly co-operative [38,39]. In contrast, folding of an RNA tetraloop with a similar loop but a much more extensive complement of interactions was shown to obey direct coupling and be non-co-operative [40]. Apparently, an extensive set of interactions

leads to additivity of mutations, while a minimal set leads to non-additivity.

Future studies are required to link co-operativity and protonation in an effort to identify motifs with highly shifted  $pK_a$  values that might participate in ribozyme catalysis. Studies on model oligonucleotides may provide the means to calculate linkage between folding and protonation in detail. Phosphorothioate NMR techniques recently described by our laboratory group [37] allow for facile determination of  $pK_a$  values and should make such investigations feasible. Previous NMR investigations tracked  $^{13}\text{C}$  chemical shifts as a function of pH and suggested that the  $pK_a$  of C75 is not highly shifted [41]. However, complications from linkage between protonation and denaturation [35] and the potential importance of the scissile phosphate in shifting, indicate the need for re-examination of this  $pK_a$ .

### Implications and perspectives

Participation of the nucleobases in chemistry significantly increases the number of catalytic devices available to RNA. Demonstration of proton transfer by the nucleobases can best be described as a histidine-like function, while demonstration of neutrality of class I  $pK_a$  values, with a potential for higher values, can best be described as a lysine-like function. Such  $pK_a$  values allow for a possible electrostatic catalysis. Several research groups have successfully increased the functional diversity of nucleic acids using organic chemistry [42,43]; although elegant, these approaches have no clear implications for the biology of extant life. Increases in the functional diversity of nucleic acids by physical chemistry (i.e. folding) discussed in the present paper are possible both in extant and emergent life. Since catalytic diversity often correlates with molecular diversity [44], large  $pK_a$  shifts may have been particularly important in a pre-protein world devoid of extensive functionality.

### References

- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) *Cell* (Cambridge, Mass.) **31**, 147–157
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) *Cell* (Cambridge, Mass.) **35**, 849–857
- Doherty, E.A. and Doudna, J.A. (2001) *Annu. Rev. Biophys. Biomol. Struct.* **30**, 457–475
- Doudna, J.A. and Cech, T.R. (2002) *Nature* (London) **418**, 222–228
- Bevilacqua, P.C., Brown, T.S., Nakano, S. and Yajima, R. (2004) *Biopolymers* **73**, 90–109
- Shan, S., Kravchuk, A.V., Piccirilli, J.A. and Herschlag, D. (2001) *Biochemistry* **40**, 5161–5171
- Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. and Strobel, S.A. (2004) *Nature* (London) **430**, 45–50
- Guo, F., Gooding, A.R. and Cech, T.R. (2004) *Mol. Cell* **16**, 351–362
- Golden, B.L., Kim, H. and Chase, E. (2005) *Nat. Struct. Mol. Biol.* **12**, 82–89
- Nesbitt, S., Hegg, L.A. and Fedor, M.J. (1997) *Chem. Biol.* **4**, 619–630
- Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M. and Scott, W.G. (1998) *Chem. Biol.* **5**, 587–595
- Nakano, S., Chadalavada, D.M. and Bevilacqua, P.C. (2000) *Science* **287**, 1493–1497
- Nakano, S., Proctor, D.J. and Bevilacqua, P.C. (2001) *Biochemistry* **40**, 12022–12038
- Karayannis, P. (1998) *Rev. Med. Virol.* **8**, 13–24
- Been, M.D. and Wickham, G.S. (1997) *Eur. J. Biochem.* **247**, 741–753

- 16 Ferré-D'Amaré, A.R., Zhou, K. and Doudna, J.A. (1998) *Nature* (London) **395**, 567–574
- 17 Wadkins, T.S., Perrotta, A.T., Ferré-D'Amaré, A.R., Doudna, J.A. and Been, M.D. (1999) *RNA* **5**, 720–727
- 18 Perrotta, A.T., Shih, I. and Been, M.D. (1999) *Science* **286**, 123–126
- 19 Nakano, S. and Bevilacqua, P.C. (2001) *J. Am. Chem. Soc.* **123**, 11333–11334
- 20 Shih, I.H. and Been, M.D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1489–1494
- 21 Bevilacqua, P.C. (2003) *Biochemistry* **42**, 2259–2265
- 22 Ke, A., Zhou, K., Ding, F., Cate, J.H. and Doudna, J.A. (2004) *Nature* (London) **429**, 201–205
- 23 Kuzmin, Y.I., Da Costa, C.P. and Fedor, M.J. (2004) *J. Mol. Biol.* **340**, 233–251
- 24 Rupert, P.B. and Ferré-D'Amaré, A.R. (2001) *Nature* (London) **410**, 780–786
- 25 Rupert, P.B., Massey, A.P., Sigurdsson, S.T. and Ferré-D'Amaré, A.R. (2002) *Science* **298**, 1421–1424
- 26 Thompson, J.E. and Raines, R.T. (1994) *J. Am. Chem. Soc.* **116**, 5467–5468
- 27 Treiber, D.K. and Williamson, J.R. (1999) *Curr. Opin. Struct. Biol.* **9**, 339–345
- 28 Chadalavada, D.M., Knudsen, S.M., Nakano, S. and Bevilacqua, P.C. (2000) *J. Mol. Biol.* **301**, 349–367
- 29 Diegelman-Parente, A. and Bevilacqua, P.C. (2002) *J. Mol. Biol.* **324**, 1–16
- 30 Chadalavada, D.M., Senchak, S.E. and Bevilacqua, P.C. (2002) *J. Mol. Biol.* **317**, 559–575
- 31 Brown, T.S., Chadalavada, D.M. and Bevilacqua, P.C. (2004) *J. Mol. Biol.* **341**, 695–712
- 32 Isambert, H. and Siggia, E.D. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6515–6520
- 33 Zamel, R., Poon, A., Jaikaran, D., Andersen, A., Olive, J., De Abreu, D. and Collins, R.A. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1467–1472
- 34 Canny, M.D., Jucker, F.M., Kellogg, E., Khvorova, A., Jayasena, S.D. and Pardi, A. (2004) *J. Am. Chem. Soc.* **126**, 10848–10849
- 35 Moody, E.M., Lecomte, J.T. and Bevilacqua, P.C. (2005) *RNA* **11**, 157–172
- 36 Misra, V.K. and Draper, D.E. (2002) *J. Mol. Biol.* **317**, 507–521
- 37 Moody, E.M., Brown, T.S. and Bevilacqua, P.C. (2004) *J. Am. Chem. Soc.* **126**, 10200–10201
- 38 Moody, E.M. and Bevilacqua, P.C. (2003) *J. Am. Chem. Soc.* **125**, 16285–16293
- 39 Moody, E.M. and Bevilacqua, P.C. (2004) *J. Am. Chem. Soc.* **126**, 9570–9577
- 40 Moody, E.M., Feerrar, J.C. and Bevilacqua, P.C. (2004) *Biochemistry* **43**, 7992–7998
- 41 Luptak, A., Ferré-D'Amaré, A.R., Zhou, K., Zilm, K.W. and Doudna, J.A. (2001) *J. Am. Chem. Soc.* **123**, 8447–8452
- 42 Piccirilli, J.A., Krauch, T., Moroney, S.E. and Benner, S.A. (1990) *Nature* (London) **343**, 33–37
- 43 Tae, E.L., Wu, Y., Xia, G., Schultz, P.G. and Romesberg, F.E. (2001) *J. Am. Chem. Soc.* **123**, 7439–7440
- 44 Carothers, J.M., Oestreich, S.C., Davis, J.H. and Szostak, J.W. (2004) *J. Am. Chem. Soc.* **126**, 5130–5137

---

Received 18 March 2005